INSTITUT FÜR INFORMATIK

Surrogate-Based Optimization of Climate Model Parameters Using Response Correction

M. Prieß, S. Koziel, T. Slawig

Bericht Nr. 1104
March 30, 2011

CHRISTIAN-ALBRECHTS-UNIVERSITÄT

ZU KIEL

Institut für Informatik der
Christian-Albrechts-Universität zu Kiel
Olshausenstr. 40
D – 24098 Kiel

Surrogate-Based Optimization of Climate Model Parameters Using Response Correction

M. Prieß, S. Koziel, T. Slawig

Bericht Nr. 1104
March 30, 2011

e-mail: mpr@informatik.uni-kiel.de, koziel@ru.is,
ts@informatik.uni-kiel.de

Surrogate-Based Optimization of Climate Model Parameters Using Response Correction

M. Prieß[a,1,*], S. Koziel[b], T. Slawig[a]

[a] *Institute for Computer Science, Cluster The Future Ocean, Christian-Albrechts Universität zu Kiel, 24098 Kiel, Germany*
[b] *Engineering Optimization & Modeling Center, School of Science and Engineering, Reykjavik University, Menntavegur 1, 101 Reykjavik, Iceland*

Abstract

We present a computationally efficient methodology for the optimization of climate model parameters applied to a (one-dimensional) representative of a class of marine ecosystem models. We use a response correction technique to create a surrogate from a temporarily coarser discretized physics-based low-fidelity model. We demonstrate that replacing the direct parameter optimization of the high-fidelity ecosystem model by iteratively updating and re-optimizing the surrogate leads to a very satisfactory solution while yielding significant cost saving - about 84% when compared to the direct high-fidelity model optimization.

Keywords: Climate models, marine ecosystem models, surrogate-based optimization, parameter optimization, response correction

1. Introduction

In this paper we present the application of a *Surrogate-based Optimization* approach, based on a multiplicative response correction, on parameter identification problems in a climate model.

Surrogate-based optimization [1–4] is a methodology to efficiently optimize complex, so-called *high-fidelity* models, that require substantial computational effort already for a model evaluation. High-fidelity models are typically evaluated through computer simulation and evaluation times of several hours, days or even weeks are not uncommon. As a consequence, optimization and control problems for such models are often still beyond the capability of modern numerical algorithms and computer power. The idea of surrogate-based optimization is to replace the high-fidelity in focus by a computationally cheaper and yet reasonably accurate representation, so-called surrogate. The surrogate can be created by approximating sampled high-fidelity model data or by employing a physically-based *low-fidelity* or coarse model. In this work, we use the latter approach. The coarse model is normally less accurate, therefore, it has to be iteratively corrected by suitable methods. The correction (or alignment) can be realized using a limited number (in many cases, only one)

*Corresponding author (*phone:* +49-(0)431 880 7452, *fax:* +49-(0)431 880 7618)
Email addresses: `mpr@informatik.uni-kiel.de` (M. Prieß), `koziel@ru.is` (S. Koziel), `ts@informatik.uni-kiel.de` (T. Slawig)
URL: `http://www.informatik.uni-kiel.de/co2/mitarbeiterinnen/dipl-phys-malte-priess/` (M. Prieß), `http://koziel.ru.is` (S. Koziel), `http://www.informatik.uni-kiel.de/co2/mitarbeiterinnen/prof-dr-thomas-slawig/` (T. Slawig)

[1] Research supported by DFG Cluster The Future Ocean

evaluations of the high-fidelity model and possibly also its derivatives. Surrogate-based optimization is widely and very successfully used in engineering sciences, compare [1–4]. The application on parameter optimization in climate models is rather new.

Climate models are typically given as time-dependent partial differential or differential algebraic equations (PDEs/DAEs) [5–7]. Since the number of processes that have to be included and the needed temporal and spatial resolution is quite high, so is the computational effort. As a result, many processes on small temporal or spatial scales are, as denoted in the climate community, *parameterized*, i.e., they are represented by simpler models that usually include a number of parameters that have to be properly chosen or adjusted. A typical example – not only used in climate models for ocean or atmosphere simulations – is turbulence modeling [8]. There are also processes in the climate system where even without much simplification several quantities or parameters are unknown or very difficult to measure. This is for example the case for growth and dying rates in marine ecosystem models [9, 10], one of which is taken as a test case for the surrogate-based optimization approach we analyze in this paper. Marine ecosystem models describe photosynthesis and other biogeochemical processes in the marine ecosystem that are important, e.g., to compute and predict the oceanic uptake of carbon dioxide (CO_2) as part of the global carbon cycle [9].

The aim of parameter optimization is to adjust or identify the model parameters so that the model output fits given measurement data [11]. The mathematical task thus can be classified as a least-squares type optimization or inverse problem [12]. The number of optimization parameters range from about 10 to 100 discrete real-valued ones in marine ecosystem models (where they are growth and dying rates etc.) up to distributed functions (or thousands and more discrete values after discretization), for example when an initial model state or boundary condition is unknown and target of the optimization. The optimization parameters and the model state are coupled by the constraint of the time-dependent PDE, i.e., the climate model. Additionally, constraints on the parameters (e.g., non-negativity of growth-rates in ecosystem models etc.) and on the state variables (non-negativity of concentrations of biological species as algae etc. or of temperature) might be given.

This optimization process requires a substantial number of (typically expensive) function and optionally sensitivity or gradient or even Hessian matrix evaluations. If the latter are computed by finite difference approximations, the critical quantity determining the computational effort of the optimization is that of the cost function evaluation, which is basically a single model simulation. Hence, decreasing the effort related to the function evaluations (or, equivalently, cutting down the number of function calls necessary to find the optimum) is of primary importance to reduce the overall optimization cost. This becomes particularly significant for computationally expensive three-dimensional coupled models, as for example global climate models [7].

In this paper we analyze the application of a *multiplicative response correction technique* to create a surrogate for one specific type of a climate model, a one-dimensional marine ecosystem model that uses pre-computed ocean circulation data [13]. This model was chosen because here extensive optimization runs with different methods including local, gradient-based and so-called global, genetic algorithms have been performed, see [14]. The underlying physically-based low-fidelity model is obtained from a temporarily coarser discretization of the high-fidelity one. We verify our approach by using synthetic target data and by comparing the results of surrogate-based optimization to those obtained from the direct fine model optimization. The application on real data is performed as a next step. Furthermore, this exemplary application shall serve as a test for three-dimensional model runs, which are much more costly with respect to computing time.

The structure of the paper is as follows: The general form of climate models and the parameter optimization problem considered is described in Section 2. We point out that the mathematical formulation of the climate models we use is quite general, such that our approach is not limited to them but remains applicable for a wide range of time-dependent models. We first recall the basic idea of surrogate-based optimization in Section 3. The ecosystem model, which is taken as an example in this paper, is introduced in Section 4, and its low-fidelity counterpart that we use as a basis for the surrogate is described in Section 5. The response correction, the construction of the surrogate model and the quality of the surrogate are described and analyzed in Section 6. The setup of the optimization which is used to compare the results is given in Section 7. Numerical results and discussion of an exemplary test run are provided in Section 8. Section 9 concludes the paper with a summary and an outlook.

2. Model Equations and Optimization Problem

In this section we give the formulations of what we call a *model* and of the corresponding parameter optimization problem. Our formulations are quite general and appropriate for a big class of applications, for which climate models are only one example.

2.1. Continuous and discrete Model Formulation

We start from an initial boundary value problem (IBVP) for a system of time-dependent partial differential or differential algebraic equations (PDEs/DAEs) of the following form:

$$\left.\begin{array}{rcll} E\,\dfrac{\partial y}{\partial t} & = & f(y,u) & \text{in } \Omega \times (0,T) \\ y(x,0) & = & y_{init}(x) & \text{in } \Omega \\ y(x,t) & = & y_{bdr}(x,t) & \text{on } \partial\Omega \times (0,T). \end{array}\right\} \qquad (1)$$

Here y is the vector of the *state variables*, and E is a matrix with the size of y, typically being the identity matrix for a PDE while having rank deficiency for a DAE [15]. We include DAEs in this formulation since in climate models, e.g., ocean circulation models, the Navier-Stokes equations [16] are an important part, and – after space discretization – take the form of a DAE system. Then y may for example consist of velocity field, pressure, temperature and salinity. In our example of a marine ecosystem model (which is formulated as PDE system), the matrix E can be set to the identity and thus omitted. In this case the state vector y contains all relevant biogeochemical tracers as phyto- and zooplankton etc., see Section 4 for the details.